Modeling Diagnostic Reasoning: A Summary of Parsimonious Covering Theory

James A. Reggia*+ and Yun Peng+

*Dept. of Neurology, Univ. of MD Hospital, Baltimore, MD 21201, USA
+Dept. of Computer Science, Univ. of MD, College Park, MD 20742, USA

ABSTRACT

Parsimonious covering theory is a formal model of diagnostic reasoning. Diagnostic knowledge is represented in the theory as a network of causal associations, and problem-solving is represented in algorithms that support a hypothesize-and-test inference process. This paper summarizes in informal terms the basic ideas in parsimonious covering theory.

INTRODUCTION

This chapter summarizes recent work on a formal model of diagnostic reasoning referred to as **parsimonious covering theory**, and discusses its relevance to medical diagnosis and expert systems. Parsimonious covering theory represents diagnostic knowledge as a network of causal associations, and models diagnostic reasoning as a hypothesize-and-test procedure whose goal is to account for observed symptoms with a plausible explanatory hypothesis. This theory, which is still evolving, captures several important features of human diagnostic inference, directly addresses the issues of diagnostic context and of multiple simultaneous causative disorders, and provides a conceptual framework within which to view recent work on diagnostic problem-solving in general.

DIAGNOSTIC REASONING AND COMPUTATIONAL MODELS OF DIAGNOSTIC INFERENCE

Diagnostic reasoning in medicine has received a great deal of attention over the last few years by cognitive psychologists, artificial intelligence (AI) researchers interested in modeling the underlying thought processes, and educators interested in improving the training of diagnosticians.[2,6,14,17,20,24,41,46] This section first briefly reviews past empirical studies of the diagnostic reasoning process, and then discusses computational models of this process. The goals are to provide the reader with sufficient background so that the relationship of parsimonious covering theory to diagnostic reasoning can be appreciated, and to relate this model to previous work in AI.

Empirical Studies of Diagnostic Reasoning

While a variety of experimental designs have been used in empirical studies of diagnostic reasoning, perhaps the most common has been the use of simulated diagnostic problems. A problem-solving session is typically recorded or videotaped, and is followed by "debriefing" of the diagnostician ("Why did you ask this? What were you thinking of here? . . ."). Based on these studies, diagnostic reasoning is generally accepted to be a sequential hypothesize-and-test (hypothetico-deductive) process during which the diagnostician conceptually constructs a "model" of the underlying causative disorders. This model, or hypothesis, is based primarily on the manifestations that are known to be present (the patient's symptoms). It postulates the presence of one or more disorders that could explain the given manifestations. Each cycle of the inference process can be viewed as consisting of three phases: disorder evocation, hypothesis evaluation or construction, and question generation. In reality, these three phases overlap extensively.

Disorder evocation is the retrieval from long-term memory of causative disorders as the diagnostician detects a new manifestation in the information available about a problem. This evoking of potential causes for the manifestations begins very early in the diagnostic process and draws on the diagnostician's memory of causal associations between disorders and their

manifestations. Ideally, the diagnostician's knowledge base or long-term memory includes the set of all possible causative disorders for each manifestation, and the set of all possible manifestations for each disorder.[46] Usually a single manifestation (rather than combinations of manifestations) is responsible for evoking new disorders for incorporation into the evolving hypothesis.[17]

The second phase of the hypothesize-and-test cycle, hypothesis evaluation or construction, involves the incorporation of possible causes of the new manifestation into the hypothesis. This may require attributing the manifestation to some disorder already assumed to be present, or adding new disorders evoked by the manifestation to the hypothesis. The diagnostician's hypothesis may at times be relatively complex. Not only may it contain a great deal of uncertainty about which of several diagnoses account for a certain manifestation, but it might also presume the simultaneous presence of multiple disorders. The empirical evidence suggests that the hypothesis can best be viewed as a resolution of two conflicting goals:

Coverage goal: The goal of explaining all of the manifestations that are present.

Parsimony goal: The goal of minimizing the complexity of the explanation.

The second goal is sometimes referred to as "Occam's razor." The parsimony goal can be viewed as an attempt to focus the reasoning process and therefore restrict searching, as a reflection of human memory limitations, or as a "common sense" heuristic that is correct most of the time.

It is important to appreciate the sequential nature of diagnostic reasoning. As the diagnositican gradually learns information about a problem, his or her hypothesis changes to reflect this new information. For example, if a patient complains of sudden onset of chest pain, a physician's initial hypothesis might be something like:

HYPOTHESIS H1: Heart attack, or pulmonary embolus, or . . .

As further details become available, some of the initially possible disorders might be eliminated. If it was then learned that the patient also had a chronic cough, the hypothesis might change to

HYPOTHESIS H2: Heart attack, or pulmonary embolus, or . . .
and
"bronchitis, or asthma, or . . ."

reflecting the physician's belief that at least two diseases must be present to account for this patient's symptoms. Note that at this point, the hypothesis contains both uncertainty (indicated by or) and the presumption that multiple simultaneous disorders are present (indicated by and).

Another aspect of hypothesis evaluation is the ranking of the likelihood of competing disorders. The term "competing disorders" refers to hypothesized alternatives which can account for the same or similar manifestations, such as heart attack and pulmonary embolus in Hypothesis H1. Perhaps surprisingly, human diagnosticians appear to use only a three-point weighting scheme to rank competing disorders: a particular finding may be "positive, noncontributory or negative with respect to a particular hypothesis."[14] At the end of a problem-solving session, diagnositicans are thus able to rank competing disorders only in a very coarse fashion (e.g., disorder d is definitely present, d is very likely to be present, d may be present, d is possible but improbable). Most of this ranking can be accounted for by either of two rules: (a) weighting based on counting the number of positive findings, or (b) weighting based on counting the number of positive findings minus the number of expected findings found to be absent.[14]

The third phase of the hypothesize-and-test cycle is question generation, and it represents the "test" phase. The word "question" here is being used in a general sense to indicate not only verbal questions, but also any type of information-gathering activity. Investigators studying human diagnostic problem-solving often divide such questions into two categories: protocol-driven and hypothesis-driven. Protocol-driven questions are those a diagnostician generally asks as a routine during a diagnostic session. In contrast, hypothesis-driven questions seek information that is specifically needed to modify the evolving hypothesis. Investigators who observe diagnosticians sometimes attribute each hypothesis-driven question to a specific

problem-solving strategy: attempting to confirm a hypothesis, attempting to eliminate a hypothesis, or attempting to discriminate between two or more hypotheses.

Many aspects of the diagnostic reasoning process are incompletely understood at the present time. For example, it is unclear how a diagnositician reasons about multiple simultaneous disorders. In such situations the manifestations must be attributed to appropriate disorders, and competing disorders must be ranked in the context of other disorders assumed to be present. It is also unclear exactly how diagnosticians decide to terminate the diagnostic process because a "solution" has been reached.

Computational Models of Diagnostic Problem-Solving

A great variety of approaches has been taken in representing and processing knowledge in computational models of diagnostic problem-solving.[31,38] Table 1 lists three prominent examples of such methods. In systems using statistical pattern classification, the knowledge base typically consists of tables of probabilities, and the inference mechanism involves the calculation of posterior probabilities of disorders using formulas such as Bayes' theorem. Models of this type have clearly achieved expert-level performance, at times outperforming human diagnosticians,[11,48] but are of limited value in situations where multiple, simultaneous disorders may occur, and are difficult to use for many real-world problems because the necessary probabilities are not available. It is important to what follows to note that models using statistical pattern classification have a strong theoretical foundation in probability theory.

Method	Theoretical Basis
Statistical pattern classification	Probability Theory
Rule-based deduction	First-order predicate calculus
Association-based abduction	?

Table 1: Methods for Diagnostic Inference.

Computational models of diagnostic problem-solving using rule-based deduction typically have a knowledge base consisting of conditional rules and an inference mechanism based on making logical deductions[19] (e.g., modus ponens or proof by refutation). As with statistical pattern classification models, systems of this type have clearly been demonstrated to exhibit an expert level of performance in empirical testing,[18,34] but reformulating naturally-occurring knowledge as rules has proven to be extremely difficult in general diagnostic domains. Rule-based, deductive systems also have a strong theoretical foundation, in this case in first-order predicate calculus.

The third approach to building diagnostic expert systems listed in Table 1 is association-based abduction. In contrast to deductive rule-based systems, whose inferences might in their simplest form be characterized by the syllogism

Given fact "A" and rule "A $\longrightarrow$ B", infer "B",

systems of this type inherently involve **abductive inference** of the form

Given fact "B" and association "A $\longrightarrow$ B", infer "plausible A".

Although the "$\longrightarrow$" in the deductive syllogism refers to logical implication, in the abductive syllogism as used in diagnostic problem-solving it refers to a causal association between A and B: "disorder A is capable of causing manifestation B, and manifestation B is known to be present, so perhaps disorder A is causing it."

The term "abduction" refers to any reasoning in which the goal is to derive the best explanation(s) for a given set of observed facts.[8,28,35,45] Tasks that are abductive in nature typically involve probabilistic, context-sensitive disambiguation of problem features using associative knowledge. They involve the **construction** of a solution, as well as selection from alternative solutions. Diagnosis is just one example of an abductive task. Many aspects of natural language understanding and high-level scene interpretation can also be viewed as abductive.[8,35]

Abductive models of diagnostic problem-solving [12,16,21,23,29,30,33,40,

[43] not only aim for a high level of performance, but they also are often explicit attempts to model the underlying reasoning of the diagnostician. Information in the knowledge base, typically causal associations, is generally represented in a descriptive or "object-oriented" fashion, and a sequential hypothesize-and-test inference process is used. The key point here for what follows is that association-based abductive models of diagnostic inference, in marked contrast to models using statistical pattern classification and rule-based deduction, do not have a readily identifiable, well-developed theoretical foundation. It is in large part due to this absence of a theoretical basis that AI research on abductive expert systems is sometimes dismissed as "ad hoc".[4] The parsimonious covering theory described in this paper represents an attempt to fill this gap.

PARSIMONIOUS COVERING THEORY

Our approach to providing a theoretical model of human diagnostic reasoning, and of association-based abductive computational systems, has been to develop a formal abductive logic referred to as **parsiminious covering theory**.[33,36] This non-deductive theoretical model has been used as the basis of a number of implemented expert systems,[33] and has proven quite powerful: it supports a descriptive knowledge representation and answer justification,[37] and it handles many of the difficulties that arise in the context of multiple simultaneous disorders.

This section describes the method used to represent diagnostic problem-solving knowledge in parsimonious covering theory and the formulation of problems. Algorithmic models of the reasoning processes used during diagnostic problem-solving are then outlined for certain classes of diagnostic problems. Finally, several issues involved in extending the basic theory are discussed. While it would be possible to present this material in a mathematically rigorous fashion, the emphasis in the following is on explaining the intuitions behind parsimonious covering theory. Readers interested in the mathematical details are referred to the literature cited above.

Knowledge Representation and Problem Formulation

In parsimonious covering theory, diagnostic knowledge is represented as an associative network of causal relationships (see Figure 1a). Disorders, indicated by nodes in set D, are causally-related to intermediate pathological states (set S), and ultimately to measurable manifestations (set M). For example, in medicine, "heart attack" would be a disorder, "shock" a pathological state, and "confused" a manifestation. A heart attack may cause shock, which in turn may cause someone to be confused. The state of being confused is considered to be a directly-observable abnormality, making it a manifestation, while shock and heart attack are not considered to be directly observable (their presence must be inferred).

The associative knowledge used in medical diagnostic problem-solving is very large and complex. One approach to formalizing the structure and use of this knowledge is to examine the most general case possible. While this is obviously the ultimate goal, as a starting point in explaining parsimonious covering it introduces a large amount of complexity and detail that obscures the central ideas of the theory. Thus, we will initially consider only the simplest version of parsimonious covering theory and discuss

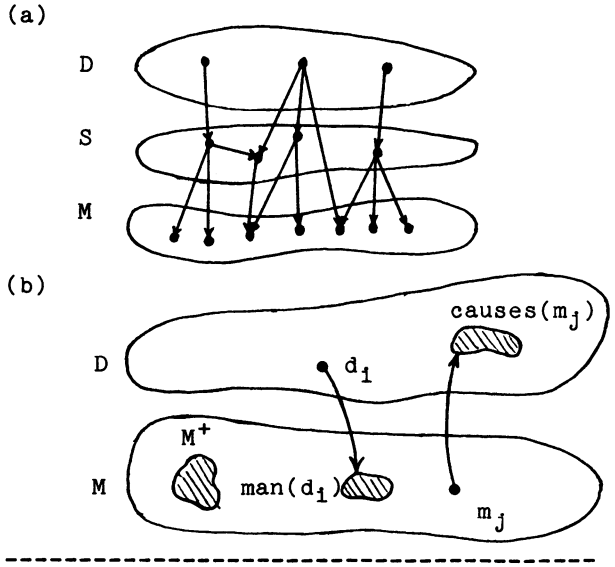

Figure 1: Causal associations in diagnostic problems.

its implications. Once this basic formulation is understood, the basic assumptions will be reexamined to explain why more general formulations are useful.

In the simplest version of parsimonious covering theory the underlying knowledge for a diagnostic problem is organized as pictured in Figure 2b. There are two discrete finite sets which define the scope of diagnostic problems: D, representing all possible disorders d_i that can occur, and M, representing all possible manifestations m_j that may occur when one or more disorders are present. In medicine D represents all known diseases (or some relevant subset of all diseases), and M represents all possible symptoms, examination findings, and abnormal laboratory results that can be caused by diseases in D. We will assume that D and M have no elements in common, and that the presence of any d_i is not directly measurable.

To capture the intuitive notion of causation, we assume knowledge of a relation C involving individual disorders and manifestations. Stating that "$\langle d_i, m_j \rangle$ is in C" represents the fact that d_i can cause m_j. Note that having $\langle d_i, m_j \rangle$ in C does not imply that m_j always occurs when d_i is present, but only that m_j may occur. For example, a patient with a heart attack may have chest pain, numbness in the left arm, loss of consciousness, or any of several other symptoms, but none of these symptoms is necessarily present.

Given D, M, and C, the following sets can be defined:

$man(d_i) = \{m_j \mid \langle d_i, m_j \rangle \text{ is in C}\}$
for each disorder d_i in D, and
$causes(m_j) = \{d_i \mid \langle d_i, m_j \rangle \text{ is in C}\}$
for each manifestation m_j in M.

These sets are depicted in Fig. 2b, and represent all possible manifestations caused by d_i, and all possible disorders that cause m_j, respectively. These concepts are intuitively familiar to the human diagnositician. For example, medical textbooks frequently have descriptions of diseases which include, among other facts, the set $man(d_i)$ for each disease d_i. Physicians often refer to the "differential diagnosis" of a symptom, which corresponds to the set $causes(m_j)$. Clearly, if $man(d_i)$ is known for every disorder d_i, or if $causes(m_j)$ is known for every manifestation m_j, then the causal relation C is completely determined. We will use $man(D_I) = \bigcup_{d_i \in D_I} man(d_i)$ and $causes(M_J) = \bigcup_{m_j \in M_J} causes(m_j)$ to indicate all possible manifestations of a set of disorders D_I and all possible causes of any manifestation in M_J, respectively. Finally, there is a distinguished subset M^+ of M which represents manifestations which are known to be present (see Fig. 2b). Whereas D, M, and C are general knowledge about a class of diagnostic problems, M^+ represents the manifestations occuring in a specific case.

Using this terminology, a (bipartite) **diagnostic problem** P in parsimonious covering theory is defined by specifying D, M, C and M^+. We will assume in what follows that diagnostic problems are well-formed in the sense that $man(d_i)$ and $causes(m_j)$ are always non-empty sets.

Having characterized a diagnostic problem in these terms, we now turn to defining the solution to a diagnostic problem by first introducing the concept of explanation.

An **explanation** E^+ for M^+ is a set of disorders where

(i) M^+ is a subset of $man(E^+)$, or E^+ **covers** M^+; and
(ii) E^+ is **parsimonious.**

An explanation represents a plausible diagnostic hypothesis. The definition of an explanation captures many features of what one intuitively means by "explaining a set of manifestations," and is central to parsimonious covering theory. Part (i) specifies the constraint that an explanation E^+ must be able to cause the manifestations known to be present in the case being diagnosed. In other words, an explanation E^+ must be a **cover**: a set of disorders that can account for or "cover" M^+. If one assumes that the manifestations in M do not occur spontaneously, but only when caused by disorders, then one can prove that whenever M^+ occurs it must be the case that one of its covers is also present.[22]

Part (ii) of the definition specifies that E^+ must also be "parsimonious," reflecting an intuitive principle often referred to as **Occam's Razor**: the simplest explanation is the preferable one. Thus, the real meta-deductive issue raised by the definition of an explanation is how one should go about formalizing the notion of

parsimony or simplicity. Here are some examples of definitions of parsimony that have been proposed on intuitive grounds for various diagnostic applications in the past:

(ii-a)Single disorder assumption: $|E^+|=1$;

(ii-b)Minimal cardinality: $|E^+| \leq |D^+|$ for any cover D^+ of M^+;

(ii-c)Irredundancy*: no proper subset of E^+ covers M^+; and

(ii-d)Relevancy: E^+ is a subset of causes(M^+).

Each of these parsimony criteria can be viewed as a generalization of those listed before it.

Assuming that only a single disorder can occur at one time (ii-a) may be appropriate in some applications,[33,43] but is obviously too restrictive for general diagnostic problem-solving. Minimal cardinality (ii-b) says that only the smallest sets of disorders covering M^+ should be considered to be plausible hypotheses. Minimal cardinality covers have been used in a number of computational models of real-world diagnostic reasoning. However, minimal cardinality covers are currently viewed as too restrictive. For example, suppose that either a very rare disorder d_1 alone, or a combination of two very common disorders d_2 and d_3, could cover all present manifestations. If minimality is chosen as the parsimony criterion, then $\{d_1\}$ would be chosen as a viable hypothesis while $\{d_2,d_3\}$ would not, even though $\{d_2,d_3\}$ might be considered to be more plausible by a human diagnostician. Thus, in many diagnostic applications, selection of plausible hypotheses based on minimal cardinality might miss the most probable hypothesis.

On the other hand, relevancy (ii-d) requires only that any disorder in an explanation be causally related to some manifestation in M^+. Intuition suggests that such a parsimony criterion would result in consideration of a large number of obviously implausible diagnostic hypotheses. Therefore, solely on an intuitive basis, a growing number of researchers in this area have adopted irredundancy as a parsimony criteria.[12,22,27,36,40] Using this criterion, a set of disorders D^+ which can account for the given manifestations M^+ is to be considered as a plausible hypothesis if and only if no proper subset of D^+ can cover M^+. In the example in the preceding paragraph, using irredundancy as a parsimony criteria results in both sets $\{d_1\}$ and $\{d_2,d_3\}$ being considered to be explanations or plausible hypotheses.

In general diagnostic problem-solving, there are typically several sets of disorders that satisfy the definition of "explanation," and one often wishes to identify all of these plausible hypotheses. Thus, the **solution** to a diagnostic problem is defined to be the set of all explanations for M^+, or some most highly ranked subset of all explanations for M^+.

d_i	man(d_i)
d_1	m_1 m_4
d_2	m_1 m_3 m_4
d_3	m_1 m_3
d_4	m_1 m_6
d_5	m_2 m_3 m_4
d_6	m_2 m_3
d_7	m_2 m_5
d_8	m_4 m_5 m_6
d_9	m_2 m_5

Table 2: A simple knowledge base.

Example. To illustrate the above concepts, let $D = \{d_1,d_2, \ldots, d_9\}$, $M = \{m_1, \ldots, m_6\}$, and let man(d_i) be as specified in Table 2. Note that Table 2 implicitly defines the relation C. Let the observable mainfestations be $M^+ = \{m_1,m_4,m_5\}$. Note that no single disorder can cover (account for) all of M^+, but that some pairs of disorders do cover M^+. For instance, if $D_1 = \{d_1,d_7\}$ then M^+ is a subset of man(D_1), so D_1 covers M^+. Since there are no covers for M^+ of smaller cardinality than D_1, it follows that D_1 is an explanation for M^+ if minimal cardinality is used as a parsimony criterion. Careful examination of Table 2 should convince the reader that, using minimal cardinality for parsimony, the solution

*Irredundancy is sometimes called "minimality," but the former term is used here to avoid any confusion with the term "minimal cardinality."

to this diagnostic problem is

$\{\{d_1,d_7\}, \{d_1,d_8\}, \{d_1,d_9\}, \{d_2,d_7\}, \{d_2,d_8\}, \{d_2,d_9\}, \{d_3,d_8\}, \{d_4,d_8\}\}$,

the set of all explanations for M^+. Alternatively, using irredundant covers as explanations,

$\{\{d_1,d_7\}, \{d_1,d_8\}, \{d_1,d_9\}, \{d_2,d_7\}, \{d_2,d_8\}, \{d_2,d_9\}, \{d_3,d_8\}, \{d_4,d_8\}, \{d_3,d_5,d_7\}, \{d_3,d_5,d_9\}, \{d_4,d_5,d_7\}, \{d_4,d_5,d_9\}\}$

is the solution or set of all irredundant covers.

In summary, the basic formulation of parsimonious covering theory presented here represents diagnostic knowledge as an associative network where causal relations link disorders to manifestations. Specifying a diagnostic problem involves specifying this associative network and a set of manifestations M^+ which are present in a specific problem. The goal of problem-solving is then to construct a set of explanations or plausible hypotheses for the given manifestations. Each exaplanation takes the form of a set of disorders which "parsimoniously cover" or account for M^+.

Procedural Models of Diagnostic Reasoning

Given a diagnostic problem formulated within the framework of parsimonious covering theory, the next issue is how to formally model the abductive, hypothesize-and-test reasoning process of the human diagnostician. A number of provably-correct procedural models exist under a variety of assumptions.[26,27,36] In the following, we will presume that manifestations are discovered one at a time (rather than all being available initially) to illustrate the flavor of parsiminious covering procedures through an informal example.

The tentative hypothesis at any point during problem-solving is defined to be the solution for those manifestations already known to be present, assuming, perhaps falsely, that no additional manifestations will be subsequently discovered. To construct a tentative hypothesis like this, three pieces of information prove useful:

MANIFS: the set of manifestations known to be present
SCOPE: causes(MANIFS), the set of all disorders d_i for which at least one manifestation is already known to be present; and
HYPOTHESIS: the tentative solution for just those manifestations already in MANIFS; typically, HYPOTHESIS is represented as a collection of generators, and should be thought of as the current "working hypothesis".

The term "generator" used here needs further definition. Rather than representing the solution to a diagnostic problem as an explicit list of all possible explanations for M^+ or MANIFS, it is advantageous to represent the disorders involved as a collection of explanation generators. An explanation generator is a collection of sets of "competing" disorders that implicitly represent a set of explanations in the solution and can be used to generate them. If A, B and C are sets of disorders, then "A x B x C," read as "one of A and one of B and one of C", is a generator that represents all explanations of the form $\{d_A,d_B,d_C\}$, i.e., all explanations where one disorder comes from A, one from B, and one from C. Hypotheses H1 and H2 (see second page) are each generators. To illustrate further this idea, consider the example diagnostic problem presented earlier (Table 2). Assuming minimal cardinality covers are explanations, two generators are sufficient to represent the solution to that problem: $\{d_1,d_2\}$ x $\{d_7,d_8,d_9\}$ and $\{d_3,d_4\}$ x $\{d_8\}$. The second generator here implicitly represents the two explanations $\{d_3,d_8\}$ and $\{d_4,d_8\}$, while the first generator represents the other six explanations in the solution.

There are at least three advantages to representing the solution to a diagnostic problem as a set of generators. First, this is usually a more compact form of the explanations present in the solution. Second, generators are a very convenient representation for developing algorithms to process explanations sequentially. Finally, and perhaps most importantly, generators are closer to the way the human diagnostician organizes the possibilities during problem solving (e.g., the "differential diagnosis").

Using the three data structures MANIFS, SCOPE and HYPOTHESIS, a hypothesize-and-test algorithm based on parsimonious covering can perform diagnostic problem solving. The HYPOTHESIS represents the tentative or working hypothesis at any point during problem-solving. The algorithm, described informally, is:

(1) Get the next manifestation m_j and add it to MANIFS.
(2) Retrieve causes(m_j) from the knowledge base.

(3) Add causes(m_j) to SCOPE.
(4) Adjust HYPOTHESIS to accommodate m_j.
(5) Repeat this process until no further manifestations remain.

Thus, as each manifestation m_j that is present is discovered, MANIFS is updated simply by adding m_j to it. SCOPE is augmented to include any possible causes d_i of m_j which are not already contained in it (derived by taking the union of causes(m_j) and SCOPE). Finally, HYPOTHESIS is adjusted to accommodate m_j based on operations involving causes(m_j) and the sets of disorders in the existing generators. These latter operations are done such that any explanation which can no longer account for the augmented MANIFS (which now includes m_j) are eliminated, and any possible new explanations are automatically constructed.

The key step in this process is Step 4, the adjustment of the HYPOTHESIS or working hypothesis. The exact form of the algebraic operations on the HYPOTHESIS depends on which parsimony criteria is being used, but in general these operations are referred to as generator "division" and "remainder" operations. Perhaps the best way to get a basic understanding of this step is to follow a simple example. Recall the abstract knowledge base illustrated in Table 2, and consider the same diagnostic problem $M^+ = \{m_1, m_4, m_5\}$ that was used earlier. Assume that irredundant covers are explanations. The order in which information about manifestations is discovered is determined by question generation heuristics. Assuming all manifestations are found, this order does not affect the final solution. For now, suppose that the sequence of events occurring during problem-solving are ordered as listed in Table 3.

Initially, MANIFS, SCOPE and HYPOTHESIS are all empty (Ø is the empty set). When m_1 is discovered to be present, m_1 is added to MANIFS, and the new SCOPE is the union of the old SCOPE with causes(m_1). Since previously there were no generators in the HYPOTHESIS, a new generator is created, in this case consisting of causes(m_1). In the terms defined earlier, this generator represents a solution for $M^+ = \{m_1\}$. It tentatively postulates that there are four possible explanations for M^+, any one of which consists of a single disorder. The HYPOTHESIS thus asserts that "d_1 or d_2 or d_3 or d_4 is present."

Events in order of their discovery	HYPOTHESIS (generator form)
Initially	Ø
m_1 present	$\{d_1\ d_2\ d_3\ d_4\}$
m_2 and m_3 absent	$\{d_1\ d_2\ d_3\ d_4\}$
m_4 present	$\{d_1\ d_2\}$ and $\{d_3\ d_4\} \times \{d_5\ d_8\}$
m_5 present	$\{d_1\ d_2\} \times \{d_7\ d_8\ d_9\}$ and $\{d_8\} \times \{d_3\ d_4\}$ and $\{d_5\} \times \{d_3\ d_4\} \times \{d_7\ d_9\}$
m_6 absent	$\{d_1\ d_2\} \times \{d_7\ d_8\ d_9\}$ and $\{d_8\} \times \{d_3\ d_4\}$ and $\{d_5\} \times \{d_3\ d_4\} \times \{d_7\ d_9\}$

Table 3: Sequential problem-solving with parsimonious covering using irredundant covers.

The absence of m_2 and m_3 does not change this initial hypothesis. However, when m_4 is discovered to be present, MANIFS and SCOPE are augmented appropriately. A new HYPOTHESIS is developed by "dividing" the only pre-existing generator set in HYPOTHESIS by causes(m_4), which in this case corresponds to intersecting causes(m_4) with the only set of disorders in the HYPOTHESIS. A new generator $\{d_1\ d_2\}$ is the result of this "division" or intersection, while the other new generator $\{d_3\ d_4\} \times \{d_5\ d_8\}$ is constructed as a "remainder" from the division process. In other words, the second new generator is built from $\{d_3\ d_4\}$ which is the part of HYPOTHESIS not kept in the division/intersection, and $\{d_5\ d_8\}$ which is the part of causes(m_4) not kept in the division/intersection. Note that the two generators at this point represent all irredundant covers for the manifestations known to be present so far. These covers have either one or two disorders in them.

When m_5 is noted to be present, MANIFS and SCOPE are again adjusted appropriately. Similar "division" and "remainder" operations are used to create a new HYPOTHESIS representing 12

irredundant covers, each containing either two or three disorders. Since m_6 is found to be absent, the resulting three generators represent exactly the irredundant covers in the final solution (compare them with the explicit listing of all irredundant covers for $M^+ = \{m_1, m_4, m_5\}$ given in the earlier example).

FURTHER ISSUES

The previous section has attempted to present the flavor of parsimonious covering theory in an informal fashion, using a simplified model of the associative knowledge involved. Clearly, as outlined earlier, more complex associative networks can be involved (see Figure 2a), and many additional issues need to be examined. This section outlines some of these issues, and is divided into three parts. The first part addresses a number of extensions relating to the basic version of parsimonious covering theory described above. The second part examines another extension, integration with probability theory, in some detail. Finally, the third part summarizes recent work to generalize the theory in a number of ways so that it addresses more complex diagnostic problems. This material is intended both to catalog many of the practical issues involved in formally modelling real-world diagnostic problem-solving, and to outline progress on addressing these issues.

Extensions to the Basic Formalism

Using even the basic form of parsimonious covering theory described above requires that a number of issues be addressed and resolved. Two related issues are how questions should be generated during sequential problem-solving to obtain additional information, and when problem-solving should terminate. These are open research questions in AI today. Diagnostic expert systems based on parsimonious covering theory[33,39] and related non-formal models[16,21,29] have generally used a heuristic approach to question generation and termination. Questions are generated in a hypothesis-driven fashion based on which disorders are currently under active consideration. Recently, an entropy minimizing metric has been proposed for abductive reasoning models.[12] All of these methods for question-generation are limited in their utility and accuracy as a model of human problem-solving.

Another issue is how M^+ should be extracted from information describing a problem. The formulation of parsimonious covering theory in the preceding section presumes that M^+ is readily available, but in some situations obtaining M^+ is non-trivial and must be inferred from a large amount of data. For examples, in the description:

"The patient is a 31 year old diabetic female with a complaint of headaches. Her temperature is 101°F, BP is 115/75, the pulse is 72/min, the neck is supple and without bruits, and the neurological examination is unremarkable except for distal loss of proprioception in the lower extremities."

it must somehow be appreciated that M^+ consists of three manifestations (headache, fever, proprioception loss).

In general, identifying M^+ involves examining the **differences** between normal behavior and observed behavior. This differencing process is quite simple in some cases. For example, identifying fever as an element in M^+ in the above description just involves recognizing that the given temperature is outside of the normal range. In other situations, extacting M^+ from observed behavior is much more involved and requires a significant amount of inferencing itself. The problem of extracting M^+ has been examined in some detail in such non-medical applications as analysis of models of biological tree growth[44] classification of errors in sequential processes[1] and in diagnosis of faults in electronic systems.[12,40]

Another practical aspect of diagnostic problem-solving is the possible availability of partial solutions. For example, in the above patient description the fact that the patient is diabetic represents an assertion that a disorder is present, not a manifestation. Applying parsimonious covering in the face of such volunteered information requires not only that diabetes be part of any overall diagnostic hypothesis, but that those elements of M^+ presumably caused by diabetes (proprioception loss) should be handled accordingly. Algorithms which perform parsimonious covering in the context of volunteered partial solutions have been developed for fairly general classes of diagnostic problems.[27]

Finally, it should be appreciated that human diagnositicians can justify

their diagnostic hypotheses, and that automated justification of problem solutions is an important feature for acceptability of diagnostic expert systems. For this reason, basic methods for automated justification of diagnostic hypotheses formulated as parsimonious covers have been studied,[37] but much work in this area remains to be done.

Integration with Probability Theory

An alternative to determining the plausibility of a diagnostic hypothesis based on a **subjective** notion of parsimony is to **objectively** calculate its probability using formal probability theory. The difficulty with this approach in the past has been that general diagnostic problems are multimembership classification problems[4]: multiple disorders can be present simultaneously. A hypothesis $D_I = \{d_1, d_2, \ldots, d_n\}$ represents the belief that disorders d_1 and d_2 and . . . and d_n are present, and that all d_i not listed in D_I are absent. Such problems are recognized to be very difficult to handle.[4,7] Among other things, letting $N = |D|$ be the total number of possible disorders, the set of 2^N diagnostic hypotheses D_I that must be ranked in some fashion is incredibly large in most real-world applications. In medicine even constrained diagnostic problems may have $50 \leq N \leq 100$, and thus 2^{50} to 2^{100} hypotheses to consider. Furthermore, there has not been any generally accepted method to rank hypotheses D_I relative to one another in multimembership problems.

Recently we have been successful in integrating formal probability theory into the framework of parsimonious covering theory in a way that overcomes some of these past difficulties.[25,26] This is achieved as follows. In the knowledge base, a prior probability p_i is associated with each disorder d_i in D where $0 < p_i < 1$. A **causal strength** $0 < c_{ij} \leq 1$ is associated with each causal association $\langle d_i, m_j \rangle$ in C, representing how frequently d_i causes m_j. For any $\langle d_i, m_j \rangle$ not in C, c_{ij} is assumed to be zero. A very important point here is that $c_{ij} \neq P(m_j|d_i)$. The probability c_{ij}, defined as $c_{ij} = P(d_i \text{ causes } m_j | d_i)$, represents how frequently d_i causes m_j when d_i is present; the probability $P(m_j|d_i)$, which is what has been used in previous statistical diagnostic models, represents how frequently m_j occurs when d_i is present. Since typically more than one disorder is capable of causing a given manifestation m_j, $P(m_j|d_i) \geq c_{ij}$. For example, if d_i cannot cause m_j at all, $c_{ij} = 0$, but $P(m_j|d_i) \geq 0$ because some other disorder present simultaneously with d_i may cause m_j. More concretely, if d_i = "heart attack" and m_j = "left ankle pain" then $c_{ij} = 0$ because a heart attack does not cause left ankle pain. However, $P(m_j|d_i)$ is greater than zero because from time to time it is the case that a patient with a heart attack also has left ankle pain.

By introducing the notion of causal strengths, and by assuming that disorders are independent of each other, that causal strengths are invariant (whenever d_i is present, it causes m_j with the probability c_{ij} regardless of other disorders that are present), and that no manifestation can occur without being caused by some disorder, a careful analysis derives a formula for $P(D_I|M^+)$, the probability of any diagnostic hypothesis D_I given the presence of any M^+, from formal probability theory.[25,26] Here D_I denotes the event that all disorders in D_I are present and all other disorders are absent, while M^+ denotes the event that all manifestations in M^+ are present and all others absent. In particular, it can be shown that

$$P(D_I|M^+) = K(M^+) \cdot L(D_I, M^+)$$

where $K(M^+)$ is a constant for all D_I given any M^+, and $L(D_I, M^+)$, called the relative likelihood of D_I given M^+, consists of three components: a weight relfecting how likely D_I is to cause the presence of manifestations in the given M^+; a weight based on manifestations expected with D_I but which are actually absent; and a weight based on the prior probabilities of disorders in D_I. Each of these weights involves only probabilistic information related to $d_i \in D_I$ and $m_j \in M^+$ instead of the entire associative network. For this reason, $L(D_I, M^+)$ is computationally very tractable.

These results make it possible to compare the relative likelihood of any two diagnostic hypotheses D_I and D_J using

$$\frac{P(D_I|M^+)}{P(D_J|M^+)} = \frac{L(D_I, M^+)}{L(D_J, M^+)}.$$

In addition to providing a method for ranking a set of parsimonious covers identified as the solution to a diagnostic problem, there are some other immediate benefits to be derived from this result as follows.

By applying this form of Bayesian classification extended to work in the

framework of parsimonious covering theory, we have been able to examine various intuitive/subjective criteria for hypothesis plausibility in an **objective** fashion. Consistent with intuition and concepts in parsimonious covering theory, probability theory leads to the conclusion that a set of disorders must be a cover to be a plausible hypothesis (i.e., non-covers have zero probability). Further, conditions can now be stated for when various criteria of "simplicity" or parsimony are reasonable heuristics for judging plausibility based on whether or not they are guaranteed to identify the most probable hypothesis.[25] For example, minimal cardinality is only appropriate to consider when all disorders are very uncommon and of about equal probability, and causal strengths are fairly large. If some disorders are relatively much more common than others, or if some causal strengths are weak, using minimal cardinality as a heuristic to select plausible diagnostic hypotheses is inadequate. In this latter situation, typical of most real-world problems, the criterion of irredundancy may be appropriate.

Irredundancy is generally quite attractive as a plausibility criterion for diagnostic hypotheses, and the set of all irredundant covers of a set of given manifestations M^+ can be shown usually to include the most likely hypothesis. However, there are two difficulties with directly generating the set of all irredundant covers as diagnostic hypotheses. First, this set may itself be quite large in some applications, and may contain many hypotheses of very low probability. Second, and more serious, it may still miss identifying the most probable diagnostic hypothesis in some cases.[25] This latter difficulty is an insight concerning plausibility criteria that has not been previously recognized.

Fortunately, both difficulties are surmountable. A heuristic function based on $L(D_I,M^+)$ can be used to guide a heuristic search algorithm to first locate a few most likely irredundant covers for M^+. Then, a typically small amount of additional search of the "neighborhood" of each of these irredundant covers can be done to see if any relevant but redundant covers are more likely. An algorithm to do this and a proof that it is guaranteed to identify the most likely diagnostic hypothesis has been presented in detail elsewhere.[26]

There are a number of generalizations that could be made to these results concerning probability theory, and we view these as important directions for further research. Our use of Bayesian classification with a causal model assumed that disorders occur independently of one another. In some diagnostic problems this is unrealistic, so a logical extension of this work would be to generalize it to such problems. Some related work has already been done along these lines in setting bounds on the relative likelihood of disorders with Bayesian classification.[9] In addition, we have developed only one method of ranking hypotheses (Bayes' Theorem) to work in causal domains involving multiple simultaneous disorders. It may be that with suitable analysis other approaches to ranking hypotheses could also be adopted in a similar fashion (e.g., Dempster-Shafer theory.[13,42]) Some initial work along these lines with fuzzy measures has already been done.[47]

Advanced Issues and Challenges

There are a number of other generalizations or extensions that have been or can be made to parsimonious covering theory. Perhaps the most obvious is that more general associative networks such as that of Figure 2a, rather than bipartite networks as in Figure 2b, can be used. This involves the use of associative or **causal chaining**: A causes B, and B causes C, so A indirectly causes C (the first two causal associations are "chained together" to form the third, reflecting the fact that causation is a transitive relationship). It has been possible to develop provably-correct algorithms that perform parsimonious covering in fairly general situations involving causal chaining.[27]

Other work in progress is investigating several related topics:

- incorporation of classification taxonomies into parsimonious covering theory;
- extension of the theory to work with causal associations involving quantified variables;
- covering M^+ with interconnected data structures rather than simple sets;
- extension of the theory to work in situations where spatial (anatomical) relationships are important;
- modification of parsimonious covering theory so that it can be

used for non-diagnostic abductive tasks such as machine learning and natural language processing;[10] and

- integration of the theory with underlying causal mechanism models.

Many of these topics are currently being examined by developing a large, real-world diagnostic system (see below).

CONCLUSION

This paper has introduced and summarized the ideas in parsimonious covering theory. While we have discussed parsimonious covering theory in the context of medical diagnosis, it should be clear that it is much more general. For example, applications using parsimonious covering or related concepts have already been developed for problem-solving in software engineering,[3] education,[15] and electronics,[12,40,43] as well as a variety of industrial applications. Although a number of significant results have been obtained in developing parsimonious covering theory, much work remains to be done to extend the formalism and increase its generality.

A pressing need at present is to test further the concepts in parsimonious covering theory through their application in real-world systems. We are currently developing a large knowledge-based diagnostic expert system which uses parsimonious covering theory. The system, called NEUREX for Neurological Expertise, will be applied to perform neurological localization and diagnosis.[39] The purpose of the NEUREX system is to serve as a substantial real-world application that will permit the critical evaluation of concepts used in parsimonious covering theory. Parsimonious covering will not only be used to perform the application-related diagnostic problem-solving, but it will also be used to support other tasks such as natural language processing and instructional use.

ACKNOWLEDGEMENTS

Supported in part by NIH Award NS-16332 and in part by NSF Award DCR-8451430 with matching funds from Software A&E, AT&T Information Systems, and Allied.

REFERENCES

1. S. Ahuja and J. Reggia, Using abductive inferencing to derive complex error simulations for discrete sequential processes, Proc. 19th Ann. Simulation Symp., D. Kembler (ed.), 207-225 (1986).
2. H. Barrows, The diagnostic (problem solving) skill of the neurologist, Arch. Neurol. 26:273-277 (1972).
3. V. Basili and C. Ramsey, ARROWSMITH-P: A prototype expert system for software engineering management, Proc. Expert Systems in Government Symposium, Karna, K., (ed.), Mclean, VA (1985).
4. M. Ben-Bassat, R. Carlson, V. Puri, et al, Pattern-based interactive diagnosis of multiple disorders: The MEDAS system, IEEE Trans. Pattern Anal. and Machine Intelligence, 2:148-160 (1980).
5. R.R. Burton, Diagnosing bugs in a simple procedural skill, in Intelligent Tutoring Systems, D. Sleeman and J.S. Brown (eds.), Academic Press, 157-183 (1982).
6. W. Card, The diagnostic process, in Medical Computing (Abrams, Ed.), American Elsevier (1970).
7. E. Charniak, The Bayesian basis of common sense medical diagnosis, Proc. of National Conference on Art. Intell., 70-73 1983.
8. E. Charniak and D. McDermott, Introduction to Artificial Intelligence, Addison-Wesley, 1985.
9. G. Cooper, NESTOR: A Computer-Based Medical Diagnostic Aid That Intergrates Causal and Probablistic Knowledge, STAN-CS-84-1031, Dept. Comp. Sci., Stanford Univ., 1984.
10. V. Dasigi and J. Reggia, Abduction in discourse processing: a parsimonious covering model, Proc. Second Symp. Role of Language in Problem Solving, 1986.
11. F. deDombal, Computer assisted diagnosis of abdominal pain, in Advances in Medical Computing (J. Rose and J. Mitchell, Eds.), Churchill-Livingston, 1975 10-19.
12. J. deKleer and B. Williams, Diagnosing multiple faults, Proc. Nat. Conf. on Artif. Intell., 1986.
13. A. Dempster, A Generalization of Bayesian Inference, Journal of Roy. Statis. Sci. Ser. 30, 205-247 1968.
14. A. Elstein, L. Shulman, and S. Sprafka, Medical Problem Solving - An Analysis of Clinical Reasoning, Harvard U.P. (1978).
15. J. Haynes, V. Pilato, D. Malouf, Expert systems for educational decision making, Education Technology, 1986, in press.
16. J. Josephson, B. Chadrasekaran, and J. Smith, Assembling the best explanation, IEEE Workshop on Principles of Knowledge-Based Systems, Denver, CO, Dec. (1984).

17. J. Kassirer and G. Gorry, Clinical problem solving, Ann. Int. Med., 89:245 1978.
18. L. Kingsland, G. Sharp, R. Capps, et al., Testing of a criteria-based consultant system in rhuematology, Proceedings of MEDINFO-83 (J. van Bemmel, M. Ball, and O. Wigertz, Eds.), North Holland, 1983, 514.
19. C. Kulikowski, Problems in the design of knowledge bases for medical consultation, Proc. First Ann. Symp. on Comp. Applic. in Medical Care, 70-74, 1977.
20. D. Leaper, et al, Clinical diagnostic process - an analysis, British Med. J. 3:569-574 (1973).
21. R. Miller, H. Pople, and J. Myers, INTERNIST-1, An experimental computer-based diagnostic consultant for general internal medicine, NEJM 307:468-476 (1982).
22. D. Nau and J. Reggia, Relationship between deductive and abductive inference in knowledge-based diagnostic problem solving, Proc. First Intl. Workshop on Expert Database Systems, Kerschberg, L. (ed.), Oct, 1984, 500-509.
23. R. Patil, P. Szolovits and W. Schwartz, Modelling knowledge of the patient in acid-base and electrolyte disorders, in Artif. Intell. in Medicine, P. Szolovits (ed.), Westview Press, 191, 1982.
24. S. Pauker, G. Gorry, J. Kassirer, and W. Schwartz, Towards the simulation of clinical cognition, Amer. J. Med. 60:981-996 (1976).
25. Y. Peng and J. Reggia, Plausibility of diagnostic hypotheses: The nature of simplicity, Proc. Nat. Conf. on Artif. Intell. 1986.
26. Y. Peng and J. Reggia, A probabilistic causal model for diagnostic problem-solving, 1986, submitted.
27. Y. Peng and J. Reggia, Diagnostic problem-solving with causal chaining, 1986, submitted.
28. C. Peirce, Abduction and Induction, Dover, 1955.
29. H. Pople, On the mechanization of abductive logic, IJCAI, 1973, 147.
30. H. Pople, Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics, in Artificial Intelligence in Medicine (P. Szolovits, Ed.), Westview Press, 1982, 119-190.
31. C. Ramsey, J. Reggia, D. Nau and A. Ferrentino, A comparative analysis of methods for expert systems, Int. J. of Man-Machine Stud., 24, 1986.
32. J. Reggia, A production rule system for neurological localization, in Proceedings of the Second Annual Symposium on Computer Applications in Med. Care, IEEE 1978, 254-260.
33. J. Reggia, D. Nau, and P. Wang, Diagnostic expert systems based on a set covering model, Internat. J. Man-Machine Studies, 19:437, 1983.
34. J. Reggia, D. Tabb, T. Price, et al., Computer-aided assessment of transient ischemic attacks, Arch. Neurol., 1984.
35. J. Reggia, Abductive inference, Proc. of Expert Systems in Government Symposium, K. Karna (ed.), 484-489 (1985).
36. J. Reggia, D. Nau and P. Wang, A formal model of diagnostic inference, Inf. Sci., 37:227 1985.
37. J. Reggia, B. Perricone, D. Nau and Y. Peng, Answer justification in diagnostic expert systems, IEEE Trans. on Biomedical Engineering, 32:263-272 (1985).
38. J. Reggia and S. Tuhrim, An overview of methods for computer-assisted medical decision making, in Computer-Assisted Medical Decision Making, J. Reggia and S. Tuhrim (eds.), Springer-Verlag, 3-45 (1985).
39. J. Reggia, S. Tuhrim, S. Ahuja, et al, Plausible reasoning during neurological problem-solving, Proc. Fifth MEDINFO, 1986, in press.
40. R. Reiter, A theory of diagnosis from first principles, TR-187/86, Dept. of Computer Science, University of Toronto, Dec. 1985.
41. A. Rubins, The role of hypotheses in medical diagnosis, in Fourth IJCAI, (1975), 856-862.
42. G. Shafer, A Mathematical Theory of Evidence, Princeton Univ., 1976.
43. H. Shubin and J. Ulrich, IDT: An intelligent diagnostic tool, in Proc. of the Natl. Conference on Artificial Intelligence, 1982, 290.
44. M. Tagamets and J. Reggia, Abductive hypothesis formation and refinement during construction of natural system models, TR-1463, Dept. of Computer Science, University of Maryland (1985).
45. P. Thagard, The best explanation - criteria for theory choice, Journal of Philosophy, 75:76-92 (1978).
46. P. Wortman, Representation and strategy in diagnostic problem solving, Human Factors 8:48, 1966.
47. R. Yager, Explanatory models in expert systems, Int. Journal of Man-Machine Studies, 23:539, 1985.
48. R. Zagoria and J. Reggia, Transferability of medical decision support systems based on Bayesian classification, Med. Decision Making, 3:501-510 (1983).